CORPUS DEL ESPAÑOL EN LOS ESTADOS UNIDOS

Manual de consulta

https://corpus.corpeeu.org

F. Javier Pueyo Mena
Francisco Gago-Jover

Con la colaboración del
Heidelberg Center for Ibero-American Studies

Versión 1.0
Mayo 2023

Tabla de contenidos

A. Interfaz de consulta....4

1 Buscador (consulta básica) ①....4

2 Generador de consultas (consulta avanzada) ②....5

3 Filtros ③....5

4 Resultados ④....7

B. Tipos de consulta....9

a. forma ortográfica....9

b. lema....9

c. etiqueta gramatical....9

d. término 1 o término 2....9

e. término 1 y término 2....9

f. término 1 y término 2 con operador de repetición....10

g. término 1a o 1b y término 2a o 2b o 2c o 2d con operador de repetición....10

C. Filtrado y ordenación de resultados....11

D. Cómo citar el CORPEEU....12

Apéndice A: Breve introducción a la sintaxis CQP (Corpus Query Processor)....13

1 Búsqueda por lemas....13

2 Búsqueda por palabras....13

3 Búsqueda por PoS....14

4 Búsqueda por frecuencia....14

5 Caracteres comodín y operadores....14

6 Búsquedas combinadas....16

6.1 Secuencial o yuxtapuesta....16

6.2 Copulativa (&)....16

6.3 Disyuntiva (|)....16

6.4 Negativa (&!)....17

7 Búsquedas complejas....17

8 Ejemplos de consultas....17

8.1 Léxico disponible en un autor, obra, o tipología....17

8.2 Consultas por marca estructural....18

8.3 Extracción de todas las formas del imperfecto de subjuntivo....18

8.4 Extracción de colocaciones del tipo “venirle en voluntad”....18

8.5 Extracción de colocaciones del tipo “verbo+pronombre_sujeto / pronombre_sujeto+verbo”....19

8.6 Extracción de casos de desdoblamiento de género “niños y niñas”, “los niños y las niñas”....20

8.7 Extracción de verbos+pronombre_enclítico (quiéranlos, tenerlos, etc.).....20
8.8 Extracción de colocaciones del tipo "sustantivo precedido por cualquier palabra excepto las de una categoría gramatical concreta".....20

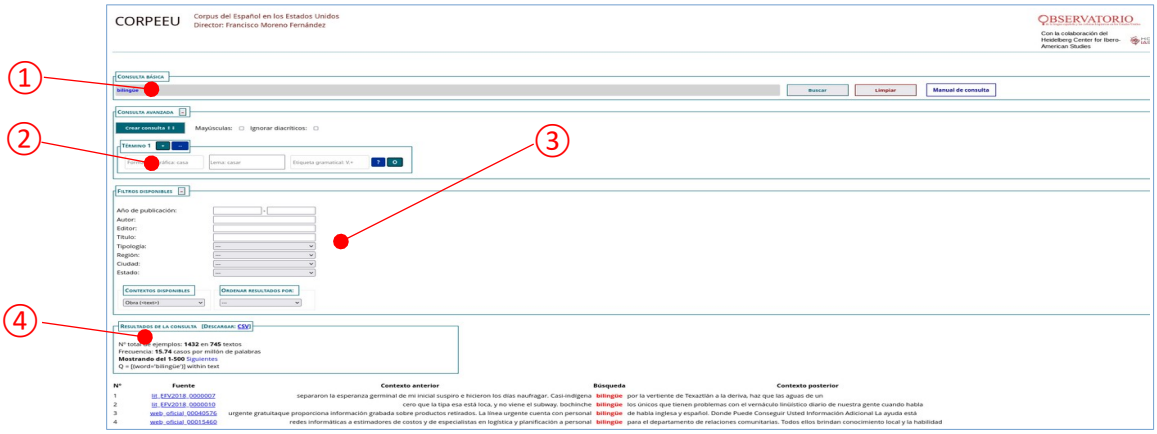
Apéndice B: Etiquetas EAGLES.....22

Apéndice C: Marcas estructurales.....30

A. Interfaz de consulta

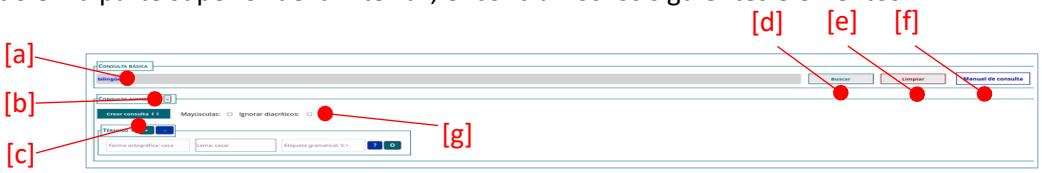
La interfaz de consulta del *Corpus del Español en los Estados Unidos* tiene cuatro componentes principales:

1 Buscador
2 Generador de consultas
3 Filtros
4 Resultados



1 Buscador (consulta básica) ①

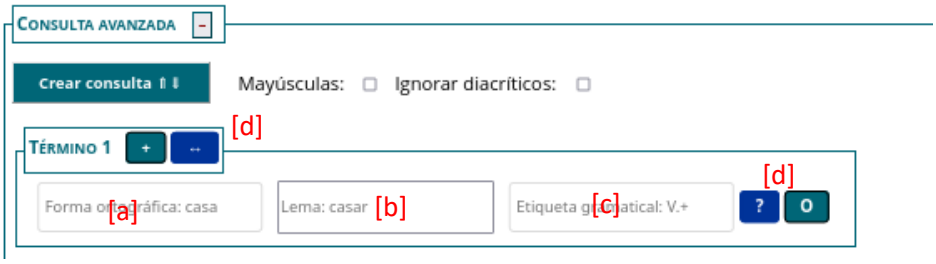
Situado en la parte superior de la interfaz, encontramos los siguientes elementos:



a Casilla de entrada de datos: los términos de la consulta pueden escribirse directamente en esta casilla (ver el ***Apéndice A*** para una explicación detallada de la sintaxis de las búsquedas) o bien pueden ser generados de forma semi-automática con el **generador de consultas** ②.
b Botón para ocultar o mostrar las opciones avanzadas de generación de consultas.
c Botón para crear e insertar la consulta
d Botón para lanzar la consulta
e Botón para limpiar la casilla de entrada de datos
f Botón para descargar el *Manual de consulta*
g Casillas de selección de mayúsculas y diacríticos

2 Generador de consultas (consulta avanzada) ②

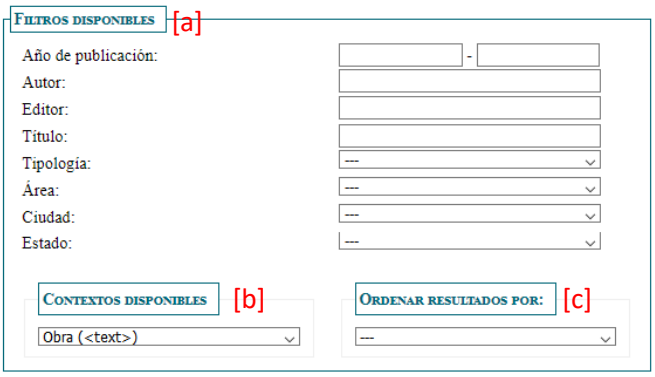
Encontramos los siguientes elementos:



a Forma ortográfica
b Lema
c Etiqueta gramatical
d Botones
- + Añadir término
- ↔ Añadir operador de repetición
- ? Definir etiqueta
- O Añadir alternativa

3 Filtros ③

En la sección correspondiente a los filtros encontramos una serie de menús desplegables que permiten el filtrado y ordenación de los resultados.



a Filtros disponibles
- Año de publicación: introducir el rango de años
- Autor: menú auto-completable
- Editor: menú auto-completable
- Título: menú auto-completable
- Tipología: menú desplegable

- Área: menú desplegable
- Ciudad: menú desplegable
- Estado: menú desplegable

b Contextos disponibles

- Marcas estructurales: restringe la consulta al texto que aparece dentro de la etiqueta estructural seleccionada, correspondiente a la etiqueta XML definida en el CORPEEU.

c Ordenación de los resultados por

- Año original de publicación
- Región
- Ciudad
- Estado
- Autor
- Editorial
- Tipología
- Forma ortográfica
- Lema
- Contexto anterior
- Contexto posterior

4 Resultados ④

Una vez efectuada la consulta los resultados se muestran en la parte inferior de la interfaz, donde encontramos los siguientes elementos

RESULTADOS DE LA CONSULTA [DESCARGAR: CSV] [a]

Nº total de ejemplos: **32705** en **14120** textos
Frecuencia: **359,40** casos por millón de palabras [b]
Mostrando del 1-500 Siguientes
Q = "casa" within text

[c]

Nº	Fuente	Contexto anterior	Búsqueda	Contexto posterior
1	lit_EFP2018_0000001	La	casa	en Mango Street No siempre hemos vivido en Mango Street. Antes vivimos en el
2	lit_EFP2018_0000001	Mamá, Papá, Carlos, Kiki, mi hermana Nenny y yo. La	casa	de Mango Street es nuestra y no tenemos que pagarle renta a nadie, ni
3	lit_EFP2018_0000001	hay propietario que golpee el techo con una escoba. Pero aún así no es la	casa	que hubiéramos querido. Tuvimos que salir volados del departamento de Loomis. Los tubos
4	lit_EFP2018_0000001	. Los tubos de agua se rompían y el casero no los reparaba porque la	casa	era muy vieja. Salimos corriendo. Teníamos que usar el baño del vecino y
5	lit_EFP2018_0000001	agua en botes lecheros de un galón. Por eso Mamá y Papá buscaron una	casa	, y por eso nos cambiamos a la de Mango Street, muy lejos,
6	lit_EFP2018_0000001	, del otro lado de la ciudad. Siempre decían que algún día nos mudaríamos a una	casa	, una casa de verdad, que fuera nuestra para siempre, de la que
7	lit_EFP2018_0000001	la ciudad. Siempre decían que algún día nos mudaríamos a una casa, una	casa	de verdad, que fuera nuestra para siempre, de la que no tuviéramos que
8	lit_EFP2018_0000001	para siempre, de la que no tuviéramos que salir cada año, y nuestra	casa	tendría agua corriente y tubos que sirvieran. Y escaleras interiores propias, como las
9	lit_EFP2018_0000001	para no tener que avisarle a todo mundo cada vez que nos bañáramos. Nuestra	casa	sería blanca, rodeada de árboles, un jardín enorme y el pasto creciendo sin
10	lit_EFP2018_0000001	árboles, un jardín enorme y el pasto creciendo sin cerca. Ésa es la	casa	de la que hablaba Papá cuando tenía un billete de lotería y ésa es la
11	lit_EFP2018_0000001	de la que hablaba Papá cuando tenía un billete de lotería y ésa es la	casa	que Mamá soñaba en los cuentos que nos contaba antes de dormir. Pero la casa
12	lit_EFP2018_0000001	casa que Mamá soñaba en los cuentos que nos contaba antes de dormir. Pero la	casa	de Mango Street no es de ningún modo como ellos la contaron. Es pequeña
13	lit_EFP2018_0000001	, y que luce todavía más chiquito entre los edificios de los lados. Nuestra	casa	tiene escaleras pero son ordinarias, de pasillo y tiene solamente un baño. Todos
14	lit_EFP2018_0000001	allí. Moví la cabeza asintiendo. Desde ese momento supe que debía tener una	casa	. Una que pudiera señalar. Pero no esta casa. La casa de Mango
15	lit_EFP2018_0000001	supe que debía tener una casa. Una que pudiera señalar. Pero no esta	casa	. La casa de Mango Street no. Por mientras dice Mamá. Es temporario
16	lit_EFP2018_0000001	tener una casa. Una que pudiera señalar. Pero no esta casa. La	casa	de Mango Street no. Por mientras dice Mamá. Es temporario, dice Papá
17	lit_EFP2018_0000001	su universo y nosotras en el nuestro. Por ejemplo mis hermanos, adentro de la	casa	tienen mucho que decirnos a mí y a Nenny. Pero afuera nadie debe verlos
18	lit_EFP2018_0000001	que es más feo que el mío. Magdalena, que por lo menos puede llegar a	casa	y hacerse Nenny. Pero yo soy siempre Esperanza. Me gustaría bautizarme yo misma
19	lit_EFP2018_0000001	. Gatos que van a dar la vuelta en la mesa del comedor. Su	casa	es el cielo de los gatos. Tú quieres una amiga, dice ella.
20	lit_EFP2018_0000001	a Francia a encontrar a la tatatatataraprima por parte de padre, y heredar la	casa	familiar. ¿Cómo lo sé? Ella me lo dijo. Entre tanto tienen

a Enlace para descargar los resultados en un fichero CSV: texto en Unicode UTF-8, con columnas separadas por tabuladores, la columna correspondiente al término buscado incluye dicho término, el <lema> y la [etiqueta gramatical]

caballos <caballo> [NCMP000]
hazlo <hacer·lo> [VMM02S0·PP3CNA00]
tenía <tener> [VMII3S0] miedo <miedo> [NCMS000]

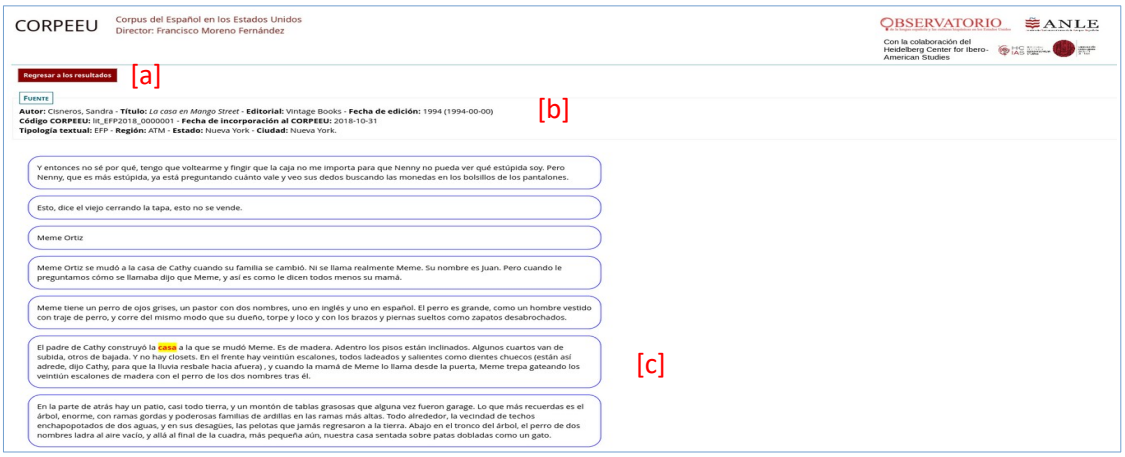
b Resultados de la consulta

- Número total de ejemplos en número de textos
- Frecuencia: casos por millón de palabras
- Indicador de resultados mostrados: permite avanzar o retroceder
- Consulta efectuada: consulta + filtros utilizados, p. ej.: `[(word=casa%c)]` :: match.text_tipologia = "EFV" & match.text_ciudad = "Nueva York" within text

c Los resultados se presentan en columnas

- Número secuencial del resultado
- Fuente: el código CORPEEU del texto que contiene el ejemplo; al pasar el ratón por encima se muestra información adicional sobre dicho texto (autoría, editorial, tipología, etc.).
- Concordancia (KWIC): Contexto anterior + Búsqueda + Contexto posterior. Al pasar el ratón por encima de cualquier palabra en la columna Búsqueda se muestra la etiqueta gramatical y el lema.

Al hacer clic en cada elemento de la columna **Fuente** se muestra el término buscado dentro de su contexto ampliado correspondiente, distinguiéndose los siguientes elementos:



a Botón para regresar a los resultados de la consulta

b Metadatos de la fuente
- i Autor
- ii Título
- iii Editorial
- iv Fecha de edición
- v Código CORPEEU
- vi Fecha de incorporación al corpus
- vii Tipología textual
- viii Región de los Estados Unidos
- ix Estado
- x Ciudad

c Contexto extendido: se muestra el fragmento con el término consultado resaltado: casa.

B. Tipos de consulta

La interfaz está diseñada de manera que no es necesario estar familiarizado con la sintaxis de consulta que utiliza el CORPEEU, pudiéndose escribir una consulta básica (con comodines o sin ellos) en la casilla de entrada de datos [a] del **buscador** ①, por ejemplo la expresión `mezquin.+`, y presionar directamente el botón **Buscar** [c].

También es posible generar de forma semi-automática un elevado número de consultas mediante el **generador de consultas** ②.

a. forma ortográfica

`(word='perro'%c)]`

1 Introducir término en la caja "Forma ortográfica"
2 Hacer clic en el botón **Crear consulta CQP ⇓ ⇑**
3 Hacer clic en el botón **Buscar**

b. lema

`[(lemma='perro'%c)]`

1 Introducir término en la caja "Lema"
2 Hacer clic en el botón **Crear consulta CQP ⇓ ⇑**
3 Hacer clic en el botón **Buscar**

c. etiqueta gramatical

`[(pos='AQ.MS.'%c)]`

1 Introducir etiqueta gramatical en la caja "Etiqueta gramatical"
2 Si se desconoce la etiqueta, hacer clic en **?** para mostrar la ayuda para definir la etiqueta. Una vez definida, hacer clic en "finalizar"
3 Hacer clic en el botón **Crear consulta CQP ⇓ ⇑**
4 Hacer clic en el botón **Buscar**

d. término 1 o término 2

`[(lemma='perro'%c) | (lemma='gato'%c)]`

1 Introducir término en la caja correspondiente
2 Hacer clic en **O** para añadir el segundo término
3 Introducir segundo término en la caja correspondiente
4 Hacer clic en el botón **Crear consulta CQP ⇓ ⇑**
5 Hacer clic en el botón **Buscar**

e. término 1 y término 2

`[(lemma='casa'%c)] [(lemma='grande'%c)]`

1 Introducir término en la caja correspondiente [término 1]
2 Hacer clic en **+** para añadir el segundo término
3 Introducir término en la caja correspondiente [término 2]
4 Hacer clic en el botón **Crear consulta CQP ⇓ ⇑**
5 Hacer clic en el botón **Buscar**

f. término 1 y término 2 con operador de repetición

`[(lemma='perro'%c)] []{0,2} [(lemma='gato'%c)]`

1 Introducir término en la caja correspondiente [término 1]
2 Hacer clic en ↔ para añadir el operador de repetición
3 Modificar el operador de repetición []{0,2} (el término anterior puede aparecer repetido entre cero y dos veces)
4 Hacer clic en + para añadir el segundo término
5 Introducir término en la caja correspondiente [término 2]
6 Hacer clic en el botón Crear consulta CQP ⇓ ⇑
7 Hacer clic en el botón Buscar

g. término 1a o 1b y término 2a o 2b o 2c o 2d con operador de repetición

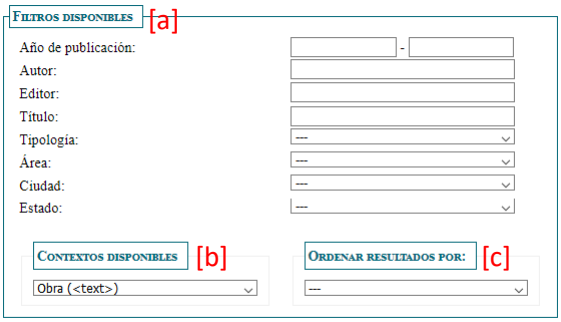
`[(lemma='haber'%c) | (lemma='tener'%c)] []{0,3} [(lemma='miedo'%c) | (lemma='vergüenza'%c) | (lemma='frío'%c) | (lemma='hambre'%c)]`

1 Introducir término en la caja correspondiente: *haber*
2 Hacer clic en O para añadir el segundo término
3 Introducir segundo término en la caja correspondiente: *tener*
4 Hacer clic en ↔ para añadir el operador de repetición
5 Modificar el operador de repetición []{0,3} (el término anterior puede aparecer repetido entre cero y tres veces)
6 Introducir término en la caja correspondiente: *miedo*
7 Hacer clic en O para añadir el segundo término
8 Introducir término en la caja correspondiente: *vergüenza*
9 Hacer clic en O para añadir el tercer término
10 Introducir término en la caja correspondiente: *frío*
11 Hacer clic en O para añadir el cuarto término
12 Introducir término en la caja correspondiente: *hambre*
13 Hacer clic en el botón Crear consulta CQP ⇓ ⇑
14 Hacer clic en el botón Buscar

Sin embargo, algunas consultas más complejas requieren que los datos sean introducidos directamente en la **casilla de entrada de datos.** En el ***Apéndice A*** se ofrece una explicación detallada de la sintaxis CQP y una serie de ejemplos concretos de diferentes tipos de consulta.

C. Filtrado y ordenación de resultados

Una vez ejecutada una consulta es posible filtrar y ordenar los resultados utilizando los "Filtros disponibles" sin necesidad de volver a recrear la consulta.



a Filtros disponibles

- Año de publicación: introducir el rango de años
- Autor: menú auto-completable
- Editor: menú auto-completable
- Título: menú auto-completable
- Tipología: menú de opciones desplegable
- Área: menú de opciones desplegable
- Ciudad: menú de opciones desplegable
- Estado: menú de opciones desplegable

b Contextos disponibles

- Marcas estructurales: restringe la consulta al texto que aparece dentro de la etiqueta estructural seleccionada.

c Ordenación de los resultados por

- Año original de publicación
- Región
- Ciudad
- Estado
- Autor
- Editorial
- Tipología
- Forma ortográfica
- Lema
- Contexto anterior
- Contexto posterior

D. Cómo citar el CORPEEU

El *Corpus del Español en los Estados Unidos*, un proyecto conjunto del *Observatorio del Español en los Estados Unidos* del *Instituto Cervantes* y de la *Academia Norteamericana de la Lengua Española* es un recurso electrónico de acceso libre y gratuito en la red con las siguientes **condiciones de uso**:

- Los usuarios citarán adecuadamente el *Corpus del Español en los Estados Unidos* en los trabajos en que utilicen datos del mismo. El formato de cita puede ser el siguiente (o uno similar que incluya los mismos datos bibliográficos):

  Francisco Moreno Fernández (dir.). 2021. *Corpus del Español en los Estados Unidos*. Observatorio del Español del Instituto Cervantes y Academia Norteamericana de la Lengua Española. En línea en https://corpus.corpeeu.org [*fecha de la consulta*]

- Las fuentes individuales deben citarse utilizando el código CORPEEU que aparece bajo la columna **Fuente** en los resultados y en los **Metadatos de fuente**.

- Para permitir que otros investigadores puedan comprobar los resultados obtenidos, se recomienda incluir la consulta efectuada tal como aparece en los **resultados** ④, incluyendo no solo el término o términos o expresión buscados, sino también los filtros empleados:

  - Q = `[(lemma='perro'%cd)]` within text
  - Q = `[(lemma='aceite'%cd)]` :: match.text_tipologia = "EFP" & match.text_ciudad = "Nueva York" within text

- Los responsables del *Corpus del Español en los Estados Unidos* agradecen la comunicación de cualquier hallazgo científico relevante que se derive de la consulta de los datos. Igualmente se ruega a los usuarios que nos informen de cualquier error textual o de cualquier fallo de funcionamiento que detecten para así contribuir a mejorar este recurso.

Apéndice A: Breve introducción a la sintaxis CQP (*Corpus Query Processor*)

El *Corpus del Español en los Estados Unidos* (CORPEEU) utiliza para las consultas la colección de herramientas de código abierto IMS Open Corpus Workbench (CWB)[1]. Su componente central es el procesador de consultas CQP (*Corpus Query Processor*) que permite realizar consultas utilizando la coincidencia de patrones con expresiones regulares. A continuación se ofrece una breve introducción a la sintaxis CQP[2].

Una de las utilidades de los corpus lematizados y etiquetados morfológicamente, como el CORPEEU, es el hecho de que se pueden efectuar consultas por atributos, es decir, por lema, palabra o categoría morfológica ("PoS" *Part of Speech*).

Lema	`[(lemma='perro')]` → *perro, perra, perros, perrita*, etc.
Palabra	`[(word='perro')]` → *perro*
PoS	`[(pos='AQ.MP.')]` → todos los *adjetivos calificativos masculino plural*

1 Búsqueda por lemas

Al efectuar una búsqueda por lemas el resultado serán todas las formas de ese lema en el corpus. Para buscar por lema, debemos utilizar la siguiente expresión `[(lemma='reina')]`. Si lo que se desea son cadenas de lemas, simplemente debemos repetir la expresión anterior tantas veces como lemas, el espacio en blanco entre ellas es opcional: `[(lemma='reina')] [(lemma='don')]`. *NOTA: No hay que dejar espacios en blanco entre las comillas y el término buscado, pues de esta forma el buscador no devolverá ningún resultado*.

Es preciso recordar que, en las consultas por lema, las mayúsculas y minúsculas son significativas y la consulta `[(lemma='alfonso')]` no devuelve ningún resultado, siendo necesario escribir el lema en mayúsculas `[(lemma='Alfonso')]` o utilizar el operador `%c` en la consulta `[(lemma='alfonso'%c)]`. Lo mismo ocurre con los diacríticos y la consulta `[(lemma='Gutierrez')]` no devuelve ningún resultado, siendo necesario poner la tilde `[(lemma='Gutiérrez')]` o utilizar el operador `%d` en la consulta `[(lemma='Gutiérrez'%d)]` para encontrar ejemplos de palabras correspondientes al lema *Gutiérrez*.

Finalmente hay que señalar que los operadores `%c` y `%d` pueden combinarse en la misma consulta: `[(lemma='gutierrez'%cd)]`

2 Búsqueda por palabras

Para buscar palabras, debemos utilizar la siguiente expresión `[(word='casa')]`. Si lo que se desea son cadenas de palabras, simplemente debemos repetir la expresión anterior tantas veces como palabras, el

[1] http://cwb.sourceforge.net/

[2] Para mayor información sobre el lenguaje de consulta CQP consúltese Evert, Stefan, et al. (2019) *The IMS Open Corpus Workbench (CWB) CQP Query Language Tutorial. CWB Version 3.4.16*. (http://cwb.sourceforge.net/files/CQP_Tutorial.pdf)

espacio en blanco entre ellas es opcional: `[(word='casa')] [(word='grande')]`. *NOTA: No hay que dejar espacios en blanco entre las comillas y el término buscado, pues de esta forma el buscador no devolverá ningún resultado.*

Es preciso recordar que, en las consultas por palabra, las mayúsculas y minúsculas son significativas y la consulta `[(word='hombre')]` solo incluye ejemplos sin mayúscula (*hombre*), siendo necesario utilizar el operador `%c` para encontrar también ejemplos con mayúscula, `[(word='hombre'%c)]` → *hombre*, *Hombre*, *HOmbre*, etc.

Lo mismo ocurre con los diacríticos y la consulta `[(word='méxi.*')]` solo incluye ejemplos de palabras *é* por (*méxico*), siendo necesario utilizar el operador `%d` para encontrar también ejemplos con *e* sin tilde `[(word='mexico.*'%d)]` → *méxico*, *mexico*, *mexicanas*, etc.

Los operadores `%c` y `%d` pueden combinarse en la misma consulta: `[(word='México'%cd)]` → *México*, *Mexico*, *Mexicanos, mexicanas,* etc.

3 Búsqueda por PoS

Para buscar categorías morfológicas (PoS), debemos utilizar la siguiente expresión `[(pos='NCFS000'%c)]`. *NOTA: La etiqueta morfológica debe ser siempre escrita en mayúscula*. Si lo que se desea son cadenas de palabras, simplemente debemos repetir la expresión anterior tantas veces como palabras, el espacio en blanco entre ellas es opcional: `[(pos='NCFS000'%c)] [(pos='AQ0FS0'%c)]`. *NOTA: No hay que dejar espacios en blanco entre las comillas y el término buscado, pues de esta forma el buscador no devolverá ningún resultado. NOTA: En el **Apéndice C** se ofrece la lista completa de las etiquetas EAGLES utilizadas en el CORPEEU.*

4 Búsqueda por frecuencia

Es posible buscar palabras o expresiones por la frecuencia (exacta, mínima o máxima) con la que aparecen en el corpus. Para encontrar todas las palabras que comienzan por 'p' en el corpus y que solo ocurren una vez debemos utilizar la expresión `[(word='p.+'%c) & (f(word) = 1)]`. Para encontrar todos los lemas que ocurren menos de 5 veces debemos utilizar la expresión `[(lemma='.+'%c) & (f(lemma) < 5)]`. Para encontrar todas las palabras que terminan en 'orio' en el corpus y que ocurren más de 2000 veces debemos utilizar la expresión `[(word='.+orio'%c) & (f(word) > 2000)]`. Finalmente, puede establecerse un rango de frecuencia mínima y máxima con la expresión `[(word='p.+'%c) & (f(word) > 100 & f(word) < 1000)]`.

5 Caracteres comodín y operadores

Supongamos ahora que queremos encontrar todas las formas del sustantivo "planta"; para ello podríamos hacer las siguientes dos consultas: `[(word='planta'%c)]` y `[(word='plantas'%c)]`, sin embargo, además de no tener los resultados en una única consulta, vemos que estos incluyen tanto la forma verbal como el sustantivo. Para evitar esto podemos utilizar caracteres comodín y operadores y efectuar una búsqueda combinada para especificar tanto formas en singular y plural como la categoría morfológica que queremos: `[(word='plantas?' & pos='NC.*'%c)]`, y lo que ahora buscamos son las

palabras *planta* o *plantas* cuando son un sustantivo (con el comodín `?` indicamos que el carácter que le precede puede aparecer cero veces o una vez, y con el operador `.*` indicamos que la etiqueta PoS debe empezar por NC). A continuación presentamos los caracteres comodín y los operadores más habituales:

`.` equivale a un único carácter: `[(word='ca.a'%c)]` → *cada*, *cala*, *cama*, *cara*, *casa*, …; `[(word='ca..a'%c)]` → *cabra*, *caída*, *carta*, *causa*, …

`.*` equivale a cero o más caracteres: `[(word='amarga.*'%c)]` → *amarga*, *amargad*, *amargado*, *amargamente*, …

`.+` equivale a uno o más caracteres: `[(word='amarga.+'%c)]` → *amargad*, *amargado*, *amargamente*, … pero no *amarga*

`..+` equivale a tres o más caracteres: `[(word='amarga..+'%c)]` → *amargados*, *amargamente* … pero no *amarga*, *amargar*, *amargado*

`?` el caracter que le precede puede aparecer cero veces o una vez: `[(word='res?pond?er'%c)]` → *responder* y *reponer*

`*` el carácter que le precede puede aparecer cero o más veces: `[(word='cas*as*'%c)]` → *casa*, *casas*, *cassa*, *cassas*

`+` el carácter que le precede puede aparecer una o más veces: `[(word='car+o'%c)]` → *caro* y *carro*

`{n}` el carácter o la expresión que le precede puede aparecer tantas veces como indique n (nº de ocurrencias): `[(word='.*o{2}.*'%c)]` → para todas las palabras que contienen dos oes seguidas; `[(pos='NP.*'%c)]{2}` → para obtener todas las ocurrencias de dos nombres propios seguidos: *Juan Pedro*, *Luis García*, etc.

`{m,n}` operador de repetición (m = número mínimo, n = número máximo de palabras): `[(pos='DA.*'%c)] [(pos='A.*'%c)]{1,2} [(lemma='hombre'%c) | (lemma='mujer'%c)]` → para obtener todas las ocurrencias de un artículo definido seguido por uno o dos adjetivos, más cualquiera de las formas de los lemas hombre o mujer.

`[]` los corchetes, sin espacio entre ellos, equivalen a una única palabra: `[(word='yo'%c)] [] [(word='doy'%c)]` → para encontrar las palabras que aparecen entre *yo* y *doy*; `[(lemma='dar'%c)] [] [(pos='SP.*'%c)] []` → para encontrar todas las palabras que ocurren en una cadena de palabras.

`[ ]` cualquiera de los caracteres dentro de los corchetes puede aparecer como mucho una vez: `[(word='pro[bv]ar'%c)]` → *probar*, *provar*.

`[ ]*` cualquiera de los caracteres dentro de los corchetes puede aparecer cero o más veces: `[(word='médic[oa][s]*'%c)]` → *médico*, *médicos, médica*, etc.

`[^ ]` cualquier carácter, excepto los caracteres dentro de los corchetes, puede aparecer: `[(word='a[^mt]ar'%c)]` → *asar*, *azar* pero no *amar* o *atar*

`!` operador de negación, `[(pos!='V.*'%c)]` → para encontrar todas las formas de las categorías gramaticales presentes en una obra, autor, etc., excepto las verbales

Diferentes caracteres comodín y operadores pueden combinarse para efectuar consultas más complejas. Por ejemplo:

- `[(word='h?a[bv]er'%c)]` → para buscar todas las variantes ortográficas (correctas o no) del infinitivo *haber*.

6 Búsquedas combinadas

Las búsquedas combinadas son aquellas que agrupan varios elementos en la misma consulta. Existen cuatro tipos básicos de búsqueda combinada:

6.1 Secuencial o yuxtapuesta

En este tipo de consulta combinamos dos o más elementos sin ninguna relación entre ellos aparte de la meramente secuencial. `[(lemma='hablar'%c)] [(word='mucho'%c)]` → para encontrar la secuencia *habla+mucho*; `[(lemma='hablar'%c)] [(pos='NP.*'%c)]` → para encontrar todas las formas de *hablar* seguidas por antropónimo.

6.2 Copulativa (`&`)

En este tipo de consulta buscamos los registros que **incluyan todos** los términos unidos mediante el operador `&`. `[(lemma='traer' & pos='V.II.*'%c)]` → para encontrar las formas de pretérito imperfecto de indicativo del verbo *traer*; `[(word='habla') & (lemma='hablar')]` → para encontrar la palabra *habla* cuando pertenece al lema *hablar*; `[(pos='V.*'%c) & (lemma='pro.*'%c)]` → para encontrar todos los verbos que comienzan por *pro-*.

6.3 Disyuntiva (`|`)

En este tipo de consulta buscamos los registros que **incluyan cualquiera** de los términos unidos mediante el operador `|`. `[(word='alto'%c) | (word='bajo'%c)]` → para encontrar las palabras *alto* o *bajo*.

6.4 Negativa (`&!`)

En este tipo de consulta buscamos los registros que **excluyan** el segundo término de los dos que aparecen unidos por el operador `&!`. `[(lemma="sierra"%c) &! (pos="V.*"%c)]` → para encontrar todas las formas del lema *sierra* excepto cuando su categoría morfológica es verbo. `[(word='a.*'%c) &! (pos='SP.*'%c)]` → para buscar todas las palabras que comienzan por *a* excepto cuando son preposiciones.

NOTA: El orden de los términos en la consulta no es arbitrario. Así `[(lemma='f.*' & pos='NP000G0'%c)]` → para encontrar todas las formas de los lemas que comienzan por *f* y son topónimos, mientras que `[(pos='NP000G0' & lemma='f.*'%c)]` → para buscar todas las palabras etiquetadas como topónimos que comienzan por *f*.

7 Búsquedas complejas

Las búsquedas complejas son aquellas que agrupan varias búsquedas combinadas en la misma consulta. A continuación se ofrecen algunos ejemplos:

- `[(lemma='traer'%c) & (pos='V.II.*' | pos='V.SI.*'%c)]` → para encontrar tanto las formas de pretérito imperfecto de indicativo como las de subjuntivo del verbo *traer*; esta búsqueda puede también escribirse así: `[(lemma='traer'%c) & (pos='V.[IS]I.*'%c)]`
- `[(lemma='traer'%c) &! (pos='V.[IS]I.*'%c)]` → para encontrar todas las formas del verbo *traer*, excepto las de pretérito imperfecto de indicativo y las de pretérito imperfecto de subjuntivo.
- `[(word='a.*'%c) &! (pos='SP.*' | pos='V.*'%c)]` → para encontrar todas las ocurrencias de *a*, excepto cuando se trata de una preposición o un verbo.
- `[(pos='NC.*'%c) & (word='[aeiou].*'%c)] [(pos='SP.*'%c) & (word='de'%c)] [(pos='NC.*'%c) & (word='[aeiou].*'%c)]` → para buscar dos sustantivos que comienzan con una vocal y están relacionados entre sí por la preposición "de".
- `[!(lemma='haber'%c) | (lemma='tener'%c)] [(lemma='bien'%c)]` → *bien* precedido por cualquier palabra, excepto las correspondientes a los lemas *haber* o *tener*.
- `[(lemma='bien'%c)] [!(lemma='haber'%c) | (lemma='tener'%c)]` → *bien* seguido por cualquier palabra, excepto las correspondientes a los lemas *haber* o *tener*.

8 Ejemplos de consultas

Para mostrar las posibilidades del *Corpus del Español en los Estados Unidos*, ofrecemos a continuación una serie de ejemplos concretos de consultas con los filtros utilizados. Debe recordarse que es posible descargar los resultados en un fichero CSV para su posterior manipulación con otros programas (Excel, R, etc.)

8.1 Léxico disponible en un autor, obra, o tipología

Esta consulta devuelve todo el léxico utilizado en un autor concreto, una obra, o una materia:

`[(word='.*'%c)]` :: match.text_autor = "Cisneros, Sandra" within text

`[(word='.*'%c)]` :: match.text_titulo_principal="Alfonso M. Becker: Volando en su escoba hacia la Casa Blanca" within text
`[(word='.*'%c)]` :: match.text_tipologia = "EFV" within text

8.2 Consultas por marca estructural

Los textos incluidos en el CORPEEU contienen una serie de marcas estructurales que reflejan la disposición del texto en la página. Esta consulta devuelve la primera palabra de cada párrafo:

`<p>[(word='.*'%c)]` within text

Esta consulta devuelve los turnos de habla que comienzan con el marcador discursivo "pues":

`<turno>[(word='pues'%c)]` within text

8.3 Extracción de todas las formas del imperfecto de subjuntivo

En la forma más sencilla de esta consulta podemos usar la etiqueta morfológica:

`[(pos='V.SI.*'%c)]` within text

para obtener una lista de todas las formas del imperfecto de subjuntivo, sin embargo podemos afinar aún más nuestros resultados con dos consultas que nos permitan sacar una lista de las formas tipo *amara* y otra de las formas tipo *amase*:

`[(word='.*[ea]ra.*' & pos='V.SI.*'%c)]` within text
`[(word='.*[ea]se.*' & pos='V.SI.*'%c)]` within text

8.4 Extracción de colocaciones del tipo "venirle en voluntad"

Esta consulta busca 1) cualquier forma de *venir* seguida por 2) cualquier forma de *voluntad*, encontrándose 1) y 2) separadas por una palabra (*vino a voluntad*, *venían con voluntad*, etc.):

`[(lemma='venir'%c)] [] [(lemma='voluntad'%c)]` within text

Si utilizamos el operador `?` la consulta busca 1) cualquier forma de *venir* seguida por 2) cualquier forma de *voluntad*, encontrándose 1) y 2) separadas por cero o una palabra (*venga voluntad*, *viene en voluntad*, etc.):

`[(lemma='venir'%c)] []? [(lemma='voluntad'%c)]` within text

Podemos también delimitar el número de palabras que separan 1) y 2). Para ello podemos repetir el operador `[]` tantas veces como palabras separan 1) y 2) (*vino hoy a voluntad*, *venir de buena voluntad*, etc):

```
[(lemma='venir'%c)] [] [] [(lemma='voluntad'%c)] within text
```

o bien podemos utilizar el operador de repetición `{m,m}` para determinar el número mínimo y máximo de palabras que pueden separar 1) y 2) (*vino voluntad*, *venía con voluntad*, *vinieron con mucha voluntad*, *venía de muy buena voluntad*):

```
[(lemma='venir'%c)] []{0,3} [(lemma='voluntad'%c)] within text
```

Finalmente, podemos encontrar colocaciones similares aumentando el número de términos en 1) y 2):

```
[(lemma='venir'%c) | (lemma='entrar'%c) | (lemma='caer'%c)] []{0,3}
[(lemma='deseo'%c) | (lemma='propósito'%c) | (lemma='voluntad'%c) |
(lemma='pensamiento'%c)] within text
```

8.5 Extracción de colocaciones del tipo "verbo+pronombre_sujeto / pronombre_sujeto+verbo"

La siguiente consulta busca todas las formas de 1ª persona del presente de indicativo de los verbos *ser*, *dar*, *ir*, *estar*, precedidas por cualquier forma del pronombre *yo*:

```
[(pos='PP1CSN00'%c)] [(lemma='dar' & pos='V.IP1S.'%c) | (lemma='ser' &
pos='V.IP1S.'%c) | (lemma='estar' & pos='V.IP1S.'%c) | (lemma='ir' &
pos='V.IP1S.'%c)] within text
```

Mientras que la siguiente consulta busca todas las formas de 1ª persona del presente de indicativo de los verbos *ser*, *dar*, *ir*, *estar*, seguidas por cualquier forma del pronombre *yo*:

```
[(lemma='dar' & pos='V.IP1S.'%c) | (lemma='ser' & pos='V.IP1S.'%c) |
(lemma='estar' & pos='V.IP1S.'%c) | (lemma='ir' & pos='V.IP1S.'%c)]
[(pos='PP1CSN00'%c)] within text
```

Es posible combinar ambas consultas y buscar simultáneamente todas las formas de 1ª persona del presente de indicativo de los verbos *ser*, *dar*, *ir*, *estar*, precedidas o seguidas por cualquier forma del pronombre *yo*:

```
[(pos='PP1CSN00'%c)] [(lemma='dar' & pos='V.IP1S.'%c) | (lemma='ser' &
pos='V.IP1S.'%c) | (lemma='estar' & pos='V.IP1S.'%c) | (lemma='ir' &
pos='V.IP1S.'%c)] | [(lemma='dar' & pos='V.IP1S.'%c) | (lemma='ser' &
pos='V.IP1S.'%c) | (lemma='estar' & pos='V.IP1S.'%c) | (lemma='ir' &
pos='V.IP1S.'%c)] [(pos='PP1CSN00'%c)] within text
```

8.6 Extracción de casos de desdoblamiento de género "*niños y niñas*", "*los niños y las niñas*"

Las siguientes consultas nos permiten extraer todos los pares de sustantivos masculinos y femeninos unidos por cualquier forma la conjunción copulativa "y" (*y*, *e*), con o sin artículo definido delante del sustantivo.

`[(pos='NCM.*'%c)] [(lemma='y'%c)] [(pos='NCF.*'%c)]` within text sort by word
`[(pos='DA.M.*'%c)] [(pos='NCM.*'%c)] [(lemma='y'%c)] [(pos='DA.F.*'%c)] [(pos='NCF.*'%c)]` within text

8.7 Extracción de verbos+pronombre_enclítico (*quiéranlos*, *tenerlos*, etc.)

Para buscar ciertas formas aglutinadas o los verbos+pronombre enclítico es preciso utilizar o bien la almohadilla (#) o bien el punto medio (· en los teclados que lo incluyen). Esta consulta permite encontrar todas las formas aglutinadas que no sean verbos

`[(lemma='.+#.+' & pos!='V.+'%c)]` within text

y esta otra todos los verbos con un pronombre_enclítico

`[(pos='V.+#.+'%c)]` within text

8.8 Extracción de colocaciones del tipo "sustantivo precedido por cualquier palabra excepto las de una categoría gramatical concreta".

En este caso queremos extraer las colocaciones del sustantivo *peso* precedido por cualquier palabra excepto artículo. Podemos utilizar una de estas dos expresiones para encontrar las formas del sustantivo:

`[(word='pesos*' & pos='NC.*'%c)]` within text
`[(lemma='peso')]` within text

Para eliminar todos aquellos casos en los que *peso* está precedido por un artículo, añadimos la expresión `[(pos!='DA.*'%c)]` a nuestra consulta:

`[(pos!='DA.*'%c)] [(word='pesos*' & pos='NC.*'%c)]` within text

de manera que se buscan combinaciones de dos palabras, donde la primera de ellas NO sea un artículo y la segunda sea *peso* SOLO cuando se trata de un nombre común.

Para excluir los artículos y las formas aglutinadas con artículos debemos utilizar la siguiente búsqueda:

`[(pos='.*'%c) &! (pos='DA.*'%c) &! (pos='SPS00#DA.*'%c)] [(word='pesos*' & pos='NC.*'%c)]` within text

en la que se buscan combinaciones de dos palabras, donde la primera de ellas tenga cualquier etiqueta morfológica, excepto artículo o la forma aglutinada preposición+artículo.

Apéndice B: Etiquetas EAGLES

El conjunto de etiquetas morfológicas utilizadas en el CORPEEU se basa en las etiquetas propuestas por el grupo *Expert Advisory Group on Language Engineering Standards* (EAGLES)[3]. A continuación se presentan cada una de las etiquetas definidas en una tabla y el conjunto de etiquetas utilizadas en el CORPEEU. Para cada categoría se presentan los atributos, valores y códigos que puede tomar, así como algunos ejemplos relevantes. Las tablas en la que se presentan las etiquetas tienen el siguiente aspecto:

Categoría			
Posición	*Atributo*	*Valor*	*Código*
columna 1	columna 2	columna 3	columna 4

En la *columna 1* encontramos un número que hace referencia al orden y posición en que aparecen los atributos. La *columna 2* hace referencia a los atributos, el número de los cuales varía dependiendo de la categoría. En la *columna 3* encontramos los valores que puede tomar cada atributo y, finalmente, la *columna 4* representa los códigos que se han establecido para su representación. Las etiquetas en sí sólo son los códigos (columna 4) y se sabe a qué atributo pertenecen por la posición (columna 1) en la que se encuentran. Así NCMS000, correspondiente a *nombre* (N), *común* (C), *masculino* (M), *singular* (S), solo tiene cuatro códigos, mientras que VMIS2P0, *verbo* (V), *principal* (M), *pretérito perfecto* (I), *simple* (S), *2 persona* (2), *plural* (P), tiene seis.

Adjetivos			
Posición	*Atributo*	*Valor*	*Código*
1	Categoría	Adjetivo	A
2	Tipo	Calificativo	Q
		Ordinal	O
3	Grado	Aumentativo	A
		Diminutivo	D
		Comparativo	C
		Superlativo	S
4	Género	Masculino	M
		Femenino	F
		Común	C
5	Número	Singular	S
		Plural	P
		Invariable	N
6	Función	-	0
		Participio	P

[3] http://www.ilc.cnr.it/EAGLES/browse.html

Adverbios			
Posición	*Atributo*	*Valor*	*Código*
1	Categoría	Adverbio	R
2	Tipo	General	G
		Negativo	N
		-mente	M

Conjunciones			
Posición	*Atributo*	*Valor*	*Código*
1	Categoría	Conjunción	C
2	Tipo	Coordinada	C
		Subordinada	S

Determinantes			
Posición	*Atributo*	*Valor*	*Código*
1	Categoría	Determinante	D
2	Tipo	Demostrativo	D
		Posesivo	P
		Interrogativo	T
		Exclamativo	E
		Indefinido	I
		Artículo	A
3	Persona	Primera	1
		Segunda	2
		Tercera	3
4	Género	Masculino	M
		Femenino	F
		Común	C
		Neutro	N
5	Número	Singular	S
		Plural	P
		Invariable	N
6	Poseedor	Singular	S
		Plural	P

Interjecciones			
Posición	*Atributo*	*Valor*	*Código*
1	Categoría	Interjección	I

Nombres			
Posición	*Atributo*	*Valor*	*Código*
1	Categoría	Nombre	N
2	Tipo	Común	C
		Propio	P
3	Género	Masculino	M
		Femenino	F
		Común	C
4	Número	Singular	S
		Plural	P
		Invariable	N
5-6	Clasificación semántica	Persona	P0
		Lugar	G0
		Organización	O0
		Otros	V0
7	Grado	Aumentativo	A
		Diminutivo	D

Numerales			
Posición	*Atributo*	*Valor*	*Código*
1	Categoría	Cifra	Z

Preposiciones			
Posición	*Atributo*	*Valor*	*Código*
1	Categoría	Adposición	S
2	Tipo	Preposición	P
3	Forma	Simple	S
		Contraída	C
4	Género	Masculino	M
5	Número	Singular	S

Pronombres			
Posición	*Atributo*	*Valor*	*Código*
1	Categoría	Pronombre	P
2	Tipo	Personal	P
		Demostrativo	D
		Posesivo	X
		Indefinido	I
		Interrogativo	T

		Relativo	R
		Exclamativo	E
3	Persona	Primera	1
		Segunda	2
		Tercera	3
4	Género	Masculino	M
		Femenino	F
		Común	C
		Neutro	N
5	Número	Singular	S
		Plural	P
		Impersonal Invariable	N
6	Caso	Nominativo	N
		Acusativo	A
		Dativo	D
		Oblicuo	O
7	Poseedor	Singular	S
		Plural	P
8	Politeness	Polite	P

Verbos			
Posición	*Atributo*	*Valor*	*Código*
1	Categoría	Verbo	V
2	Tipo	Principal	M
		Auxiliar	A
		Semiauxiliar	S
3	Modo	Indicativo	I
		Subjuntivo	S
		Imperativo	M
		Infinitivo	N
		Gerundio	G
		Participio	P
4	Tiempo	Presente	P
		Imperfecto	I
		Futuro	F
		Pasado	S
		Condicional	C
5	Persona	Primera	1
		Segunda	2
		Tercera	3

6	Número	Singular	S
		Plural	P
7	Género	Masculino	M
		Femenino	F

Signos de puntuación			
Posición	*Atributo*	*Valor*	*Código*
1	Categoría	Puntuación	F

ADJETIVOS

Ordinal

AO0MS0
AO0MP0
AO0FS0
AO0FP0

Calificativo

AQ0MS0
AQ0MP0
AQ0FS0
AQ0FP0
AQ0CS0
AQ0CP0
AQ0CN0

ADVERBIOS

General

RG

-mente

RM

Negativo

RN

CONJUNCIONES

Coordinada

CC

Subordinada

CS

DETERMINANTES

Artículo

DA0MS0
DA0MP0
DA0FS0
DA0FP0
DA0NS0

Demostrativo

DD0MS0
DD0MP0
DD0FS0
DD0FP0
DD0CS0
DD0CP0
DD2FS0
DD2MS0

Exclamativo

DE0MS0
DE0MP0
DE0FS0
DE0FP0
DE0CS0
DE0CP0
DE0CN0

Indefinido

DI0MS0
DI0MP0
DI0FS0
DI0FP0
DI0CS0
DI0CP0

Posesivo

DP1CSS
DP1CPS
DP2CSS
DP2CPS
DP3CS0
DP3CP0
DP1MSP
DP1MPP
DP1FSP
DP1FPP
DP2MSP
DP2MPP
DP2FSP
DP2FPP

Interrogativo

DT0MS0
DT0MP0
DT0FS0
DT0FP0
DT0CS0
DT0CP0
DT0CN0

INTERJECCIONES

I

NOMBRES

Común

NCMS000
NCMP000
NCFS000
NCFP000
NCMN000
NCFN000
NCCS000
NCCP000
NCCN000

Propio

NP00000

Topónimo

NP000G0

Antropónimo

NP000P0

Extranjerismo

NCEX000

NUMERALES

Z

PREPOSICIONES

SPS00

PRONOMBRES

General

P0000000
P01CP000
P02CP000
P03CN000

Demostrativo

PD0MS000
PD0MP000
PD0FS000
PD0FP000
PD0CS000
PD0CP000
PD0NS000

Exclamativo

PE000000
PE0MS000
PE0MP000
PE0FS000
PE0FP000
PE0CS000
PE0CP000

Indefinido

PI0MS000
PI0MP000
PI0FS000
PI0FP000
PI0CS000
PI0CP000

Personal

PP1CSN00
PP1CS000
PP1CSO00
PP1CP000
PP1MP000
PP1FP000
PP2CSN00
PP2CS000
PP2CSO00
PP2CP000
PP2CP00P
PP2CS00P
PP2MP000
PP2FP000
PP3MS000
PP3FS000
PP3MP000
PP3FP000
PP3NS000
PP3FSA00

PP3FPA00
PP3MSA00
PP3CNA00
PP3MPA00
PP3CSD00
PP3CPD00
PP3CN000
PP3CNO00

Relativo

PR000000
PR0MS000
PR0MP000
PR0FS000
PR0FP000
PR0CS000
PR0CP000
PR0CN000

Interrogativo

PT000000
PT0MS000
PT0MP000
PT0FS000
PT0FP000
PT0CS000
PT0CP000
PT0CN000

Posesivo

PX1MS0S0
PX1MP0S0
PX1NS0S0
PX1FS0S0
PX1FP0S0
PX1MS0P0
PX1MP0P0
PX1NS0P0
PX1FS0P0
PX1FP0P0
PX2MS0S0
PX2MP0S0
PX2NS0S0
PX2FS0S0
PX2FP0S0
PX2MS0P0
PX2MP0P0
PX2NS0P0
PX2FS0P0
PX2FP0P0
PX3MS000
PX3MP000
PX3FS000
PX3FP000
PX3NS000

VERBOS (AUXILIAR)

Infinitivo

VAN0000

Gerundio

VAG0000

Participio pasado

VAPS0SF
VAPS0SM
VAPS0PF
VAPS0PM

Presente de indicativo

VAIP1S0
VAIP2S0
VAIP3S0
VAIP1P0
VAIP2P0
VAIP3P0

Pretérito imperfecto

VAII1S0
VAII2S0
VAII3S0
VAII1P0
VAII2P0
VAII3P0

Pretérito perfecto simple

VAIS1S0
VAIS2S0
VAIS3S0
VAIS1P0
VAIS2P0
VAIS3P0

Futuro de indicativo

VAIF1S0
VAIF2S0
VAIF3S0
VAIF1P0
VAIF2P0
VAIF3P0

Condicional

VAIC1S0
VAIC2S0
VAIC3S0
VAIC1P0
VAIC2P0
VAIC3P0

Presente de subjuntivo

VASP1S0
VASP2S0
VASP3S0
VASP1P0
VASP2P0
VASP3P0

Pretérito imperfecto de subjuntivo

VASI1S0
VASI2S0
VASI3S0
VASI1P0
VASI2P0
VASI3P0

Futuro de subjuntivo

VASF1S0
VASF2S0
VASF3S0
VASF1P0
VASF2P0
VASF3P0

Imperativo

VAM02S0
VAM03S0
VAM01P0
VAM02P0
VAM03P0

VERBOS (PRINCIPAL)

Infinitivo

VMN0000

Gerundio

VMG0000

Participio presente

VMPP0SC
VMPP0PC

Participio pasado

VMPS0SF
VMPS0SM
VMPS0PF
VMPS0PM

Presente de indicativo

VMIP1S0
VMIP2S0
VMIP3S0
VMIP1P0
VMIP2P0
VMIP3P0

Pretérito imperfecto

VMII1S0
VMII2S0
VMII3S0
VMII1P0
VMII2P0
VMII3P0

Pretérito perfecto simple

VMIS1S0
VMIS2S0
VMIS3S0
VMIS1P0

VMIS2P0
VMIS3P0

Futuro de indicativo

VMIF1S0
VMIF2S0
VMIF3S0
VMIF1P0
VMIF2P0
VMIF3P0

Condicional

VMIC1S0
VMIC2S0
VMIC3S0
VMIC1P0
VMIC2P0
VMIC3P0

Presente de subjuntivo

VMSP1S0
VMSP2S0
VMSP3S0
VMSP1P0
VMSP2P0
VMSP3P0

Pretérito imperfecto de subjuntivo

VMSI1S0
VMSI2S0
VMSI3S0
VMSI1P0
VMSI2P0
VMSI3P0

Futuro de subjuntivo

VMSF1S0
VMSF2S0
VMSF3S0
VMSF1P0
VMSF2P0
VMSF3P0

Imperativo

VMM02S0
VMM03S0
VMM01P0
VMM02P0
VMM03P0

VERBOS (SEMIAUXILIAR)

Infinitivo

VSN0000

Gerundio

VSG0000

Participio pasado

VSPS0SF
VSPS0SM
VSPS0PF
VSPS0PM

Presente de indicativo

VSIP1S0
VSIP2S0
VSIP3S0
VSIP1P0
VSIP2P0
VSIP3P0

Pretérito imperfecto

VSII1S0
VSII2S0
VSII3S0
VSII1P0
VSII2P0
VSII3P0

Pretérito perfecto simple

VSIS1S0
VSIS2S0
VSIS3S0
VSIS1P0
VSIS2P0
VSIS3P0

Futuro de indicativo

VSIF1S0
VSIF2S0
VSIF3S0
VSIF1P0
VSIF2P0
VSIF3P0

Condicional

VSIC1S0
VSIC2S0
VSIC3S0
VSIC2P0
VSIC1P0
VSIC3P0

Presente de subjuntivo

VSSP1S0
VSSP2S0
VSSP3S0
VSSP1P0
VSSP2P0
VSSP3P0

Pretérito imperfecto de subjuntivo

VSSI1S0
VSSI2S0
VSSI3S0
VSSI1P0
VSSI2P0
VSSI3P0

Futuro de subjuntivo

VSSF1S0
VSSF2S0
VSSF3S0
VSSF2P0
VSSF1P0
VSSF3P0

Imperativo

VSM02S0
VSM03S0
VSM01P0
VSM02P0
VSM03P0

SIGNOS DE PUNTUACIÓN

F

Apéndice C: Marcas estructurales

Las etiquetas utilizadas en las fuentes para indicar la estructura del texto y la disposición de los diferentes elementos presentes en cada página son las siguientes.

Etiqueta CORPEEU[4]	Descripción
<text>	*Encabeza cada fuente textual del corpus*
<text_id>	*Identificador de la fuente textual*
<text_autor>	*Autor*
<text_titulo_principal>	*Título*
<text_editorial>	*Editorial o sitio web*
<text_tipologia>	*Tipología textual*
<text_region>	*Región de los Estados Unidos*
<text_estado>	*Estado*
<text_ciudad>	*Ciudad*
<text_fecha_original>	*Fecha de edición original*
<text_fecha_digital>	*Fecha de incorpración al corpus*
** **	*Párrafo*
<turno>	*Turno de habla (en entrevista orales y medios audiovisuales)*
<alargamiento>	*Alargamiento en la pronunciación (en entrevista orales y medios audiovisuales)*
<cita>	*Cita literal (en entrevista orales y medios audiovisuales)*
<enfasis>	*Énfasis en la pronuncación (en entrevista orales y medios audiovisuales)*
<entre_risas>	*Texto enunciado entre risas (en entrevista orales y medios audiovisuales)*
<ininteligible>	*Enunciación ininteligible (en entrevista orales y medios audiovisuales)*
<interrupcion_de_grabacion>	*Grabación interrumpida (en entrevista orales y medios audiovisuales)*
<lengua>	*En otra lengua (en entrevista orales y medios audiovisuales)*
<palabra_cortada>	*Palabra cortada (en entrevista orales y medios audiovisuales)*
<pausa>	*Pausa (en entrevista orales y medios audiovisuales)*
<pausa_type>	*Tipo de pausa (en entrevista orales y medios audiovisuales)*
<risas>	*Risa (en entrevista orales y medios audiovisuales)*
<risas_type>	*Tipo de risa (en entrevista orales y medios audiovisuales)*
<observacion_complementaria>	*Observaciones sobre eventos ocurridos durante la grabación*

[4] Cada una de estas etiquetas posee su correspondiente etiqueta de cierre: , </turno>, etc.